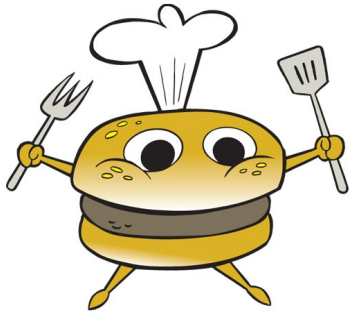
BUlgarian **R**esource **G**rammar – **E**fficient and **R**obust



Petya Osenova
BulTreebank Group
Institute for Parallel Processing
Bulgarian Academy of Sciences

DELPH-IN Summit, Paris, 2 July 2010

Overview

- Matrix-based (via customization page as of January 2010)
- Full coverage of MRS test suite
- Coverage of additional language-specific phenomena
- Incorporation of full morphology using LKB types/rules
- Documentation of theory and implementation
- Preparing to link to large Bulgarian Treebank
- Ready for inclusion as DELPH-IN resource

Some Linguistic Phenomena

- Relatively free word order
- Most arguments optional
- Rich morphological agreement
- Rich clitic system
- Idiosyncracies with auxiliaries, complementizers (of course)

Coverage of Test Suite

‘1006’ Coverage Profile							
Length	**total items** ♯	**positive items** ♯	**word string** ϕ	**lexical items** ϕ	**distinct analyses** ϕ	**total results** ♯	**overall coverage** %
10 – 14	2	2	10.00	169.50	24.00	2	100.0
5 – 9	77	73	5.49	61.44	5.16	73	100.0
1 – 4	134	119	3.15	23.47	1.92	119	100.0
Total	**213**	**194**	**4.10**	**39.26**	**3.37**	**194**	**100.0**

(generated by [incr tsdb()] at 1-jul-2010 (17:55 h))

Issues

- Revisions to Matrix types
  - Semantic index of (intersective) adjectives, adverbs
  - HEAD-ADJ-PHRASE not QUE 0-dlist, not POSTHEAD +
  - BASIC-HEAD-OPT-COMP-PHRASE also for nominal phrases
- Desire to cache lexical rule filter
  - 2000+ inflectional rules
  - Recomputing takes 2-3 minutes
  - Changes less often than lexicon
- Cyrillic font
  - Trees okay, not LKB input window
  - Probably solved if use LOGON configuration

Next Steps

- Slavic language family
  - Tania Avgustinova has created candidate `slavic.tdl`
  - Russian Matrix grammar now three-level implementation
  - Plan to factor Bulgarian types, to use `slavic.tdl`
- Bulgarian Lexicon
  - Plan to extend inventory of BURGER verb lexical types
- Bulgarian Treebank
  - 200,000 word corpus of (partly) manually constructed trees
  - Web page: `bultreebank.org`
  - Plan to define mapping from BURGER trees to BulTreeBank
- Coverage extensions